Construct validity of the Heart Failure Screening Tool (Heart-FaST) to identify heart failure patients at risk of poor self-care: Rasch analysis

DR CHANTAL SKI (Orcid ID : 0000-0003-1324-2933)

PROFESSOR DAVID ROBERT THOMPSON (Orcid ID : 0000-0001-8518-6307)

Article type : Research Methodology: Instrument Development

Construct validity of the Heart Failure Screening Tool (Heart-FaST) to identify heart failure patients at risk of poor self-care: Rasch analysis

Running head: **Heart-FaST Rasch analysis**

Nicholas A. REYNOLDS (BSc)

Research Associate, Mental Health Service, St Vincent's Hospital, Melbourne, Australia

Chantal F. SKI (PhD)

Reader, School of Nursing and Midwifery, Queen's University, Belfast, United Kingdom

Samantha M. McEVEDY (LLB)

Research Manager, School of Psychology and Public Health, La Trobe University, Melbourne, Australia

David R. THOMPSON (PhD)

Professor, School of Nursing and Midwifery, Queen's University, Belfast, United Kingdom

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jan.13544

Jan CAMERON (PhD)

Senior Research Fellow, School of Clinical Sciences, Monash University, Melbourne, Australia

Correspondence

Chantal Ski, School of Nursing and Midwifery, Queen's University, Belfast BT9 7BL, UK

Email: C.Ski@qub.ac.uk

Acknowledgements – N/A

Conflict of Interest Statement

No conflict of interest has been declared by the authors

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Aim. To psychometrically evaluate the Heart Failure Screening Tool (Heart-FaST) via: i) examination of internal construct validity; ii) testing of scale function in accordance with design; and iii) recommendation for change/s, if items are not well adjusted, to improve psychometric credential.

Background: Self-care is vital to the management of heart failure. The Heart-FaST may provide a prospective assessment of risk, regarding the likelihood that patients with heart failure will engage in self-care.

Design. Psychometric validation of the Heart-FaST using Rasch analysis.

Method: The Heart-FaST was administered to 135 patients (Median age = 68, IQR = 59-78 years; 105 males) enrolled in a multidisciplinary heart failure management program. The Heart-FaST is a nurse-administered tool for screening patients with HF at risk of poor self-care. A Rasch analysis of responses was conducted which tested data against Rasch model expectations, including whether items serve as unbiased, non-redundant indicators of risk and measure a single construct and that rating scales operate as intended.

Results: The results showed that data met Rasch model expectations after rescoring or deleting items due to poor discrimination, disordered thresholds, differential item functioning, or response dependence. There was no evidence of multidimensionality which supports the use of total scores from Heart-FaST as indicators of risk.

Conclusion: Aggregate scores from this modified screening tool rank heart failure patients according to their 'risk of poor self-care' demonstrating that the Heart-FaST items constitute a meaningful scale to identify heart failure patients at risk of poor engagement in heart failure self-care.

Key Words

heart failure, Heart-FaST, instrument development, Rasch analysis, self-care

Summary statement

Why is this research needed?

- Heart failure is one of the most common causes of hospitalisations, morbidity and mortality and requires intensive self-care.
- The complexity of heart failure self-care places a severe strain on patients, often leading to poor motivation and/or adherence.

What are the key findings?

- Based on findings from the Rasch analysis, recommendations are made (item reduction and rescoring) to improve measurement effectiveness of the Heart-FaST for the assessment of poor engagement in heart failure self-care.
- The modified Heart Failure Screening Tool (Heart-FaST) is a valid tool for identifying patients at risk of poor self-care across physical, cognitive and emotional domains.

How should the findings be used to influence policy / practice / research / education?

- Used as a screening tool to identify patients at risk of poor engagement in self-care across cognitive, emotional and physical domains, the HeartFaST has the potential to be a valuable tool to assist in the development of patient care plans
- To substantiate the association between poor engagement in HF self-care and performance on the HeartFaST, further validation in a larger sample size is warranted.

INTRODUCTION

Heart failure (HF) affects approximately 26 million people worldwide (Bui et al., 2011). This complex syndrome is often associated with periods of acute cardiac decompensation and is among the leading causes of hospitalisation, morbidity and mortality in older adults (Bui et al.; 2011; Ambrosy et al., 2014). Patients with HF must become proficient in self-care if they are to fully benefit from treatment (Cameron et al., 2012; Jonkman et al., 2016). HF self-care is a multifaceted process with expectations that patients monitor their symptoms, appraise changes that occur, promptly redress problems or seek medical assistance and remain confident in their ability to perform these tasks (Riegel et al., 2016).

Poor engagement in self-care has been shown to precipitate poor quality of life, clinical decompensation and premature mortality (Jonkman et al., 2016; Lee et al., 2016). This points to the need for a valid screening tool to identify patients at risk of poor engagement in self-care. Such a tool would enable i) systematic and prompt identification of a vulnerable patient population and ii) remedial strategies to be implemented to minimise risk of adverse health outcomes. Patients with HF who are identified as being 'at risk', for instance, may be advised to monitor their weight and fluid retention daily and report an increase in weight of 2kg over two days. Early identification of fluid retention via this simple intervention could instigate remedial efforts, potentially averting dire or lethal consequences (Lee et al., 2011). Yet poor self-care can have manifold causes and presentations, which complicates the task of identifying patients risk in that regard (Riegel et al., 2009a). While screening tools to assess HF self-care behaviours have been in use for some time (Riegel et al., 2009b; Vellone et al., 2014), they are retrospective self-reports. Only recently has a test been devised that may help nurses to identify patients who might be at risk of poor engagement in self-care (Cameron et al., 2014).

Background

The Heart Failure Screening Tool (Heart-FaST) collates information about patient attributes that can hinder engagement in self-care, including physical, cognitive and emotional limitations (Cameron et al., 2014). The Heart-FaST is designed to help clinicians tailor and apply disease-management strategies on an individual needs basis, using domain-specific scores in conjunction with cut-off points. Nursing services can thus be tailored to redress physical, emotional or cognitive needs of patients with HF or any combination of these factors (Cameron et al., 2014). However, a unitary construct may govern responses to the Heart-FaST items and therefore this scale may help to identify patients who are likely to struggle with self-care. In other words, total scores may provide a useful index of risk insofar as physical limitations (e.g. comorbidities) increase the likelihood of concomitant cognitive and emotional limitations, all of which affect a patient's capacity for self-care. In that case, increasing scores would signify an accumulation of issues that complicate self-care, via an increased burden of care or a diminished capacity in that regard.

The Rasch model has been identified as useful for rigorous examination and development of measurement instruments in nursing research (Hagquist et al., 2009). Rasch analysis is the process of testing data against a mathematical ideal of measurement (Andrich, 2004; Newby et al., 2010). In this way, Rasch analysis is a confirmatory process in that data should demonstrate their credentials with respect to measurement theory and not the other way around. A mathematical rendition of ideal measurement, the Rasch model specifies conditions that responses to a set of items will uphold if a construct has been measured (Christensen et al., 2013; Rasch, 1960, 1980; Tennant and Conaghan, 2007). Thus, for the purpose of the current study, Rasch analysis was deemed the most appropriate method of analysis to test whether i) the Heart-FaST items serve as unbiased and non-redundant indicators of risk, ii) whether rating scales operate as intended and iii) whether responses accord with the Rasch model, which characterises fundamental measurement in the physical and social sciences. Dependent on meeting these requirements, responses might then be summed to a total score across the Heart-FaST domains and interpreted as signifying variation along a single continuum (i.e. risk of poor self-care).

THE STUDY

Aims

To psychometrically evaluate the Heart Failure Screening Tool (Heart-FaST) via: i) examination of internal construct validity; ii) testing of scale function in accordance with design; and iii) recommendation for change/s, if items are not well adjusted, to improve psychometric credential.

Methodology

A psychometric evaluation of the Heart-FaST was conducted using Rasch analysis of the Heart-FaST with RUMM2030 (Andrich et al., 2009). Rasch analysis was identified as the most appropriate model as it facilitates disclosure of lack of invariance and other measurement problems that may not be easily detected by more traditional analyses. The analysis used partial credit parameterization which allows inter-threshold distances to vary across the scale's items (Masters, 1982). The empirical operation of rating scales was assessed before interpreting fit statistics and seeking evidence for multidimensionality, differential item functioning and local dependence. As recommended by Linacre (1994) for Rasch measurement, our sample size (N=135) was adequate to attain 99% confidence that no item calibration is more than 1 logit away from its stable value.

To test whether successive categories of each rating scale signify increasingly higher levels of risk, we compared the empirical order of thresholds with the order that was designated to them. If thresholds were reversed, relative to their intended order, then we imposed an order on these data. This was done by collapsing categories that manifested non-modal category probability functions (i.e. at no point along the continuum were they ever the most likely categories to be endorsed).

We used statistical and graphical indicators of fit to evaluate data against Rasch model expectations. We inferred that data fit the Rasch model provided that (a) the item-trait interaction chi-square was non-significant (α level of .05), which implies that items occupy the same relative position at any level of risk, (b) standard deviations of the distributions of standardized fit residuals for persons and items were less than 1.5, (c) individual item fit residuals were within ± 2.0 and item chi-square tests were non-significant (α level of .05, Bonferroni adjusted) and (d) graphical displays corroborated the fit statistics such that observed proportions means (plotted at three class intervals along the continuum) were aligned with their expected values (i.e. item characteristic curves).

We tested whether items operate in the same way regardless of age group (28-61, 62-74, or 75-96 years) or gender (male and female) with analysis of variance (ANOVA) of the residuals (Bonferroni adjusted α level of .05). Cut-off scores for Heart-FaST were implemented as per those recommended by Cameron et al. (2014). We used the Q_3 statistic to test whether items are related exclusively via the latent attribute (i.e. local item independence), using a critical value of .3 (Yen, 1993). Q_3 is the partial correlation between the residuals of a pair of items after the latent attribute has been accounted for (Yen, 1993). Residual correlations can indicate duplication of one item by another (i.e. response dependence) or result from factors that are exogenous to the latent attribute (i.e. trait dependence) (Marais and Andrich, 2008).

We used the person separation index (PSI) as a measure of internal consistency reliability. Calculated from the person measures (i.e. non-linear transformations of raw scores) and their error terms, the PSI is interpreted in the same way as Cronbach's alpha (Andrich, 1982). If items were misaligned with persons, which inflates the error or measurement and reduces the PSI, we calculated a sample-independent estimate of reliability using the method devised by Wright (2001). This entails advancing along the possible range of scores by twice the joint standard error of their corresponding logit measures (i.e. the square root of the sum of squared standard errors) (Wright 2001). Wright's sample independent reliability estimate indicated the maximum number of strata that a test can reliably differentiate along the score range.

Finally, we tested whether the Heart-FaST measures a unitary construct by testing the equivalence of two theoretically identical measures of risk for each person (Smith, 2002). Measures were derived from mutually exclusive subsets of items with opposite loadings on the first principal component of the residuals. Items with positive or negative loadings represent contrasting aspects of a secondary dimension (i.e. the first residual dimension after conditioning out the Rasch dimension). We inferred that the latent attribute is unidimensional if the proportion of participants with differences outside the range of ±1.96 was less than 5%, according to the lower limit of a binomial confidence interval (normal approximation) (Smith, 2002; Tennant and Conaghan, 2007).

Participants

This study included patients enrolled in a multidisciplinary HF management program from four hospitals in Melbourne, Australia. Inclusion criteria were as follows: 18 years of age or older and a confirmed diagnosis of HF (symptoms and clinical features of congestion that occur at rest or on exertion of minimal effort) with objective evidence of underlying structural cardiac dysfunction from an echocardiogram or an elevated plasma level of B-type natriuretic peptide (BNP). Exclusion criteria were as follows: residing in an aged care, high needs facility (nursing home); a documented history of moderate-to-severe cognitive impairment or dementia; a terminal diagnosis; and insufficient comprehension to read English without the need of a translator. Data collection took place between March 2012 - June 2013. Participants were approached in the cardiac wards by registered nurses who then referred them to a research assistant who explained the study in detail and obtained written consent. The Heart-FaST was administered in a face to face interview either in hospital or at the participant's home. Demographics and clinical data were captured in person and confirmed in patient's medical record. All patients who participated in the study were administered the Heart-FaST (Cameron et al., 2014).

Instrument

The Heart-FaST is a clinician-administered screening tool for the management of HF self-care (Cameron et al., 2014). It comprises 17 self-report items and nine items that are scored by the test administrator following the completion of a task. Points are accrued for failing a task or responding in the affirmative to a self-report item. In its proposed function as a unidimensional risk assessment, responses are summed to a total score that is presumed to increase with a HF patient's risk of poor self-care.

Items in the Heart-FaST represent one of three domains: cognitive functioning, emotional functioning and physical functioning and demographic risk factors. The latter domain comprises the New York Heart Association (NYHA) functional classification (Criteria Committee of the New York Heart Association, 1964) (scored 0, 1, 2 after combining NYHA classes I and II); three demographic characteristics, each scored as 1 if confirmed (being 70 years of age or older, residing alone, being a non-native English speaker); having someone to confide in (disagree scored as 1); the Charlson comorbidity index (Charlson et al., 1987) (scored 0, 1, 2 after combining Charlson categories 3, 4 and 5); and the time elapsed since receiving a diagnosis of HF (less than two months scored as 1). The cognitive functioning domain comprises four self-report items that describe difficulties with concentration and memory (scored as 1 for an affirmative response to each item) and five objectively scored items. These include two time-limited verbal fluency tasks (scored 0-2), a digit span working memory task (scored 0-2); a serial sevens task (scored 0-5); and a clock drawing task with three components: ‘Contours’ (scored 0, 1), ‘Numbers’ (scored 0-2) and ‘Hands’ (scored 0-2). Finally, the domain of emotional functioning consists of seven statements from the Cardiac Depression Scale (Hare et al., 1996) that describe one's mood and perceived prospects (e.g., "there is only misery in the future for me"). Patients respond using a seven-category rating scale, which ranges from ‘strongly agree’ to ‘strongly disagree' and enumerated 0-7. Further information about item scoring is available elsewhere (Cameron et al., 2014).

The content validity of the Heart-FaST has been established, as has the feasibility of administering it in clinical settings (Cameron et al., 2014). However, questions remain about the appropriate use of total scores for the Heart-FaST.

Two considerations support the contention that responses can be summed to a total score across the domains of functioning (Hobart and Cano, 2009). First, the Heart-FaST includes only items with some bearing on HF self-care, according to theory and expert consensus (Cameron et al., 2014). In other words, the items have a common function, as indicators of risk, though each domain assesses a distinct aspect of functioning as it pertains to self-care (Cameron et al., 2014). Second, physical, cognitive and emotional limitations may be interdependent in the context of HF, with physical limitations affecting the cognitive and emotional wellbeing of patients (Alosco et al., 2015a,b; Hawkins et al., 2015). For example, the physiological changes induced by HF can compound pre-existing issues with memory and concentration and even precipitate cognitive impairment (Cannon et al., 2015). The link between emotional distress and HF is less well qualified. However, HF can impose an immense burden of suffering on patients, of which emotional distress, including depression and anxiety, are likely manifestations (Sokoreli et al., 2016).

Ethical considerations

Ethics approval was granted by the Eastern Health Human (LR39-1112) and Australian Catholic University Ethics Committees (2012 04V). All methods conformed to the principles outlined in the Declaration of Helsinki (World Medical Association, 2013).

RESULTS

Sample characteristics

Characteristic features of the sample (N=135) are summarised in Table 1. The median age of the sample (70% male) was 68 years, with two-thirds having NYHA Class I-II and the remaining one-third Class III-IV.

Initial assessment of fit

Data did not accord with the Rasch model, two items failed to discriminate between patients as a function of risk (“Less than 2 months since diagnosis", fit residual = 2.18, χ^2 (df =123.84) = 11.93, p = 0.002; “Do you live alone?", fit residual = 2.79, non-significant χ^2). They were omitted from the scale. The subsequent analyses revealed evidence of local dependencies, differential item functioning and disordered thresholds. Items were rescored or omitted from the scale to gauge the effects of these anomalies on scaling. Table 2 reports the person separation index and summary fit statistics following each modification.

Modifications due to threshold disorder

The rating scale for six emotional functioning items failed to operate as intended. Mid-range categories were undifferentiated with respect to the latent attribute. It was reduced from seven to two categories by collapsing non-modal categories that were adjacent to one another. The first three categories were combined into a single category while another category was created from the rating scale's remaining four categories. In addition to disordered thresholds, the serial sevens rating scale did not operate as intended. Mid-range categories were undifferentiated with respect to the latent attribute. Consequently, the mid-range categories were combined to form a single, intermediate category, indicative of partial success on this task.

Modifications due to local item dependence

High residual correlations between self-report items indicated that they were related beyond their common function as indicators of risk. Self-report of "difficulty concentrating and thinking" increased the likelihood of reporting "difficulty keeping your attention on an activity for long" (Q_3 = .38) and "trouble with your memory" (Q_3 = .34). To circumvent the effect of local dependence on scaling, "difficulty concentrating and thinking" was omitted from the analysis.

High residual correlations were also present between the scoring criteria of Clock Hands and those of Clock Numbers. Residuals were correlated Q_3=.65 between Clock Numbers i (“All numbers must be present with no additional numbers”) and Clock Numbers ii (“Numbers must be in the correct order and placed in the approximate quadrants on the clock face”). Residuals were correlated Q_3=.47 between Clock Hands i (“There must be two hands jointly indicating the correct time”) and Clock Hands ii (“The hour hand must be clearly shorter than the minute hand and must be centred within the clock face with their junction close to the clock centre”). The criteria were modified by allocating one point for failing either criterion of each task.

Modifications due to differential item functioning

The likelihood of reporting concern about one's capacity for sexual activity varied as a function of gender (Figure 1). A uniform pattern of differential item functioning was seen: at any level or risk, females were less likely than males to endorse this item, $F(1, 133) = 14.13$, $p=0.0002$. Consequently, it was among the most readily endorsed items by males (δ = -0.65 logits) whereas females were very unlikely to endorse it (δ = 0.46 logits). The item was omitted from the analysis because the difference between parameters differed by >.5 logits as a function of gender, which was deemed a substantive difference.

Fit to the Rasch model following modifications

Following these modifications, data met Rasch model expectations (Table 2). Item fit residual statistics were within ±2.0 and chi-square tests were non-significant. There was no evidence of differential item functioning or local dependencies. Item loadings on the first principal component of the residuals revealed a contrast between items that are scored by the test administrator and self-report items. The difference between measures from these subtests was statistically significant for 9 HF patients. This amounted to 6.7% of the sample, with a lower confidence limit that crossed 5% (95% CI 3.1-10.3%).

Figure 2 shows the alignment of items (or thresholds) with persons along the continuum of risk. Items tended to target the upper third of the distribution of measures ($\bar{X}$=-1.23, SD ±0.81). The average participant will have found it difficult to accrue points on the average item. Items and thresholds ranged between ±2 logits while person measures ranged from -3 to +1 logits. Measurement precision was greatest around zero logits, in accord with the designated function of the Heart-FaST: to identify patients who are likely to be at risk of poor engagement in HF self-care.

Measurement error was high because most items measured 'high' risk, whereas most persons exhibited 'low' or 'intermediate' risk. However, a sample-independent estimate of reliability indicated that the reported person separation index of .61 did not indicate the true measurement effectiveness of the scale. The sample-independent reliability estimate revealed that a maximum of four statistically distinct levels of performance could be discerned along the possible score range, corresponding to a person separation index of .94.

Invariant item parameter estimates and fit statistics from the modified Heart-FaST are shown in Table 3. Having no one to confide in, failing the contours component of the clock drawing task, feeling that there is “only misery in the future,” and failing every serial sevens subtraction task were indicative of the highest risk level measured by the Heart-FaST. By contrast, items that define the lowest risk level in this context were self-report of poor memory, the Charlson comorbidity index and a sense of living on borrowed time. Figure 3 illustrates the scaling of items and HF patients in a common additive unit along the continuum of risk. It illustrates the correspondence between qualitative (i.e. item content) and quantitative indicators of risk at various levels along the continuum.

DISCUSSION

Our study has shown that the Heart-FaST data are consistent with the partial credit derivation of the Rasch model, after items were deleted or rescored due to poor discrimination, local dependencies, differential item functioning or threshold disorder. Results support the contention that risk of poor engagement in self-care is a unidimensional construct. Scores from this modified screening tool rank patients with HF by their 'risk of poor self-care' and might usefully be analysed with non-parametric statistics (Tennant and Conaghan, 2007).

We found that signs of mild cognitive impairment, isolation and despair characterise the highest level of risk measured by the Heart-FaST. Based on the putative meaning of scores, patients with HF would manifest a high risk of poor engagement in self-care who failed to construct a circle while attempting the clock drawing task; who failed every serial sevens subtraction task; who had no one to confide in and who felt that the future held only misery for them. Such patients would be considered strong candidates for remedial attention.

We found also that burdens of disease and aging characterise the lowest level of risk, including self-reported memory problems, comorbidity and a sense of living on borrowed time. Patients with HF who accrue points on these items, but not items situated above them in the hierarchy, would likely manifest a low or intermediate risk of poor engagement in self-care. However, research is needed to substantiate the validity of this interpretation; there is currently limited evidence that scores indeed do predict poor engagement in HF self-care, or the problems affiliated with it (e.g. use of medical services, unexpected hospital admissions, or poor prognosis as a function of performance on the Heart-FaST).

Finally, the Heart-FaST items appear to constitute a logical scale, when their content is considered in relation to their relative locations along the continuum of risk. With respect to items that measure the cognitive component of risk, the simplest such items occupy the scale's uppermost positions, whereas more complex faculties of cognition are required of the items situated below them in the hierarchy. It follows that failing a simple task confers a greater risk of poor engagement in HF self-care than would failing a complex task. Also logical is the sequence of items that measure the emotional component of risk: signs of extreme distress, such as despair, supersede more moderate or transient forms of distress, such as feeling in low spirits. In summary, the most likely pattern of responses to items is meaningful in that items define a logical progression from low to high risk of poor engagement in self-care.

This study identified three problems with the measurement of risk by the Heart-FaST. Rating scales failed to operate as intended, items were related beyond their common function as indicators of risk and the operation of one item varied according to the gender of the patient responding to it. Identified as statistical anomalies, these issues are aspects of the data that depart from the characteristics of ideal measurement. They likely have a substantive basis in the measurement system which ought to be investigated and understood, according to the experimental measurement paradigm to which Rasch analysis belongs (Tennant and Conaghan, 2007).

The first anomaly, known as threshold disorder, occurs when higher ratings do not consistently signify 'more' of the attribute being measured (Andrich, 2013a,b). The rating scale for items that measure the emotional component of risk originally contained seven categories. Yet mid-range categories were undifferentiated along the continuum and indicated that it operated instead as a dichotomy. It is difficult to interpret scores unless it can be shown that the integers assigned to successive categories correspond to increasing levels of the latent attribute. This was not demonstrated with respect to the emotional functioning rating scale; thus, a reduction in the number of categories is warranted. The continued use of a seven category rating scale, despite evidence that it does not function appropriately, would undermine the validity of inference based on scores (Andrich, 2013a,b). As regards the causes of threshold disorder, participants may have struggled to discern seven levels of agreement when responding to these items (Andrich and Styles 2005). The verbal administration of items will have compounded any problems arising from an excess of categories because it denies patients a visual reference by which to respond (i.e. a line with category labels).

Mid-range categories of the serial sevens rating scale were also undifferentiated with respect to the latent attribute. HF patients did not tend to fail subtraction tasks in their serial order, nor did one or two failed subtractions signify a lower risk of poor self-care than three or four failed subtractions. This is not the first report of rating scale dysfunction for the serial sevens task. Koski and colleagues (2009) found that thresholds were reversed, relative to their intended order, in the four-category version of serial sevens employed by the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). The authors recommended against modifying how serial sevens is scored because the MoCA is a well-established test (Nasreddine et al., 2005). Nevertheless, fewer than four levels of mild cognitive impairment can be discerned in the context of measuring mild cognitive impairment, while fewer than six levels of performance were evident in the present context.

Response dependence was the second anomaly identified in this study (Marais and Andrich, 2008; Wright, 1996). A specific violation of the requirement for statistical independence among responses, it can reflect duplication of one item by another, which governs determines how participants are likely to respond to it. In this study, a patient who had had difficulty concentrating and thinking was more likely to report problems with memory than would be anticipated based on his or her risk level. The item "difficulty concentrating and thinking" was omitted from the scale, as it was thought to provide little unique information beyond what was already provided by its locally dependent counterparts (Marais and Andrich, 2008).

High residual correlations between the criteria of Clock Hands and those of Clock Numbers were also considered to be symptomatic of response dependence (Marais and Andrich, 2008). The second criterion of Clock Numbers, for instance, stipulates that numbers be placed "in the correct order and placed in the approximate quadrants on the clock face". This requires that numbers are present to begin with, which is the first criterion of Clock Numbers. Thus, to accrue points on the second criterion of each task, it was necessary to have failed the first criterion. When these items were analysed as polytomies, the same problem manifested as threshold disorder. To circumvent the effects of local dependence on scaling, the first criterion of each task was omitted from the scale (Marais and Andrich, 2008). The proposed revised scoring criteria of Clock Numbers and Clock Hands now resemble those used by the MoCA (Nasreddine, et al, 2005).

Finally, men with the same level of risk as women were nevertheless more likely to report concern about sexual dysfunction. The variable operation of this item may reflect medical problems that are specific to male HF patients. For example, the effect of medication on erectile function (specifically, thiazide diuretics and most beta-blockers) (Baumhakel et al, 2011) or whatever mechanisms underpin the association between erectile dysfunction and cardiovascular disease (Gandaglia et al., 2014). The item was omitted from the scale because its location differed substantially as a function of gender (i.e. > 1 logit) and because sexual functioning was not considered essential to the measurement of risk (Scheuneman and Subhiyah, 1998).

In summary, we recommend that the Heart-FaST be modified in four ways, to improve the scale's measurement effectiveness. First, amalgamate the scoring criteria of Clock Hands and those of Clock Numbers. Second, construct three serial sevens rating scale categories by awarding two points for failing every subtraction, one point for failing one or more subtractions and zero points for successfully completing every subtraction (i.e. complete success, partial success, complete failure). Third, ask participants to respond with a dichotomous format to emotional functioning items (e.g. agree or disagree). Fourth, omit the item which asks about sexual functioning as well as two items that discriminate poorly along the continuum (i.e. "less than two months since diagnosis?" and "do you live alone?").

Research has yet to establish a link between aggregate scores from the Heart-FaST and the likelihood that patients will not adequately engage in the process of HF self-care via maintenance behaviours, symptom monitoring or symptom management, for instance (Lee et al., 2011) . Physical, cognitive and emotional limitations are known to impinge on self-care, which can render patients vulnerable to adverse health outcomes. Studies are now needed to substantiate putative link between poor engagement in HF self-care and performance on the Heart-FaST to support its intended use as a screening tool.

Limitations

These findings are preliminary, in part because rating scales were modified *post hoc*, for exploratory purposes, following evidence that rating scales do not function according to their design (Andrich, 2013a). Eliciting responses to emotional functioning items with a dichotomous format is likely to change how patients interact with them. This has the potential to alter the hierarchical order of items, relative to which scores are interpreted. Therefore, the relevance of these findings remains unclear until studies implement a rating scale with fewer than seven categories and report the effect of this modification on the measurement of risk.

The findings are further limited by the sample size, which was insufficient because most items target 'high' risk, a region of the continuum where very few patients were located. This can diminish the power of fit statistics and the precision of parameter estimates (Smith et al., 2008). Studies of the proposed revised version of the Heart-FaST will require no fewer than 250 patients to overcome this problem. A sample size of 150 is sufficient provided that measures span the range of calibrations. In other words, studies with sample sizes of less than 250 should recruit patients suspected of being vulnerable to poor engagement in self-care (Smith et al., 2008).

While the results of this study are consistent with the idea that physical, cognitive and emotional limitations converge along a single continuum, the theory of risk is not well articulated. No hypotheses were made, for instance, about the relative order of items based on an understanding of what various levels of risk might actually entail. As the potential for falsification is limited at this stage, it is difficult to assess the internal construct validity of the Heart-FaST. Future studies should also examine the association of the Heart-FaST scores with other self-care measures and health outcomes. The study has heuristic value nevertheless; it may stimulate inquiry into HF self-care, including the precipitants of poor engagement in self-care that can leave patients especially vulnerable to adverse health effects.

CONCLUSION

Poor engagement in HF self-care has manifold causes and presentations. This can complicate efforts to redress the burden of HF to individuals, the community and the health care system. The Heart-FaST collates patient characteristics that are known to impede engagement in self-care. It might thereby help identify patients at risk of poor engagement in self-care. A Rasch analysis of a modified version of the Heart-FaST has shown that data conform to modern psychometric standards. Items constitute a meaningful scale along which patients can be rank-ordered, according to their risk of poor engagement in HF self-care.

Author Contributions:

All authors have agreed on the final version and meet at least one of the following criteria (recommended by the ICMJE*):

1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;

2) drafting the article or revising it critically for important intellectual content.

* http://www.icmje.org/recommendations/

References

Alosco M.L., Spitznagel M.B., Cohen R., Sweet L.H., Hayes S.M. Josephson, R., Hughes J. & Gunstad J. (2015a) Decreases in daily physical activity predict acute decline in attention and executive function in heart failure. *Journal of Cardiac Failure* **21**(4), 339-346. doi: 10.1016/j.cardfail.2014.12.010

Alosco M.L., Spitznagel M.B., Sweet L.H., Josephson R., Hughes J. & Gunstad J. (2015b) Cognitive dysfunction mediates the effects of poor physical fitness on decreased functional independence in heart failure. *Geriatriatrics and Gerontology International* **15**(2), 174-181. doi: 10.1111/ggi.12245

Ambrosy A.P., Fonarow G.C., Butler J., Chioncel O., Greene S.J., Vaduganathan M., Nodari S., Lam C.S., Sato N., Shah A. N. & Gheorghiade M. (2014) The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *Journal of the American College of Cardiology* **63**(12)**,** 1123-1133. doi: 10.1016/j.jacc.2013.11.053

Andrich D. (1982) An index of person separation in latent trait theory, the traditional KR-20 index and the Guttman scale response pattern. *Education Research and Perspectives* **9**(1), 95-104.

Andrich D. (2004) Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care* **42**(1S), i7-i16. doi: 10.1097/01.mlr.0000103528.48582.7c

Andrich D. (2013a) An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "Threshold disorder controversy". *Educational and Psychological Measurement* **73**(1)**,** 78-124. doi: 10.1177/0013164412450877

Andrich D. (2013b) The legacies of R. A. Fisher and K. Pearson in the application of the polytomous Rasch model for assessing the empirical ordering of categories. *Educational and Psychological Measurement* **73**(4), 553-580. doi: 10.1177/0013164413477107

Andrich D. & Styles I. (2005) *Report on the Rasch analysis of the Australian Early Development Instrument (AEDI) using 2004 data from 6 states. Australian Early Development Census*. Perth, WA: Murdoch University.

Andrich D., Lyne A., Sheridan B.C. & Luo G. (2009) *RUMM2030: A windows interactive program for analysing item response data according to Rasch Unidimensional Measurement Model.* Perth, WA: RUMM Laboratory Ltd.

Baumhakel M., Schlimmer N., Kratz M., Hackett G., Jackson G. & Bohm, M. (2011) Cardiovascular risk, drugs and erectile function - a systematic analysis. *International Journal of Clinical Practice* **65**(3), 289-298. doi: 10.1111/j.1742-1241.2010.02563.x

Bui A.L., Horwich T.B. & Fonarow G.C. (2011) Epidemiology and risk profile of heart failure**.** *Nature Reviews Cardiology* **8**(1)**,** 30-41. doi: 10.1038/nrcardio.2010.165

Cameron J., Ski C.F. & Thompson D.R. (2012) Screening for determinants of self-care in patients with chronic heart failure. *Heart, Lung and Circulation* **21**(12)**,** 806-808. doi: 0.1016/j.hlc.2012.07.050

Cameron J., Ski C.F., McLennan S.N., Rendell P.G., Whitbourn R.J. & Thompson D.R. (2014) Development of the Heart Failure Screening Tool (Heart-FaST) to measure barriers that impede engagement in self-care. *European Journal of Cardiovascular Nursing* **13**(5)**,** 408-417. doi: 10.1177/1474515113502461

Cannon J., McMurray J. & Quinn T. (2015) 'Hearts and minds': association, causation and implication of cognitive impairment in heart failure. *Alzheimers Research and Therapy* **7**(1), 22. doi: 10.1186/s13195-015-0106-5

Charlson M.E., Pompei P., Ales K.L. & MacKenzie C.R. (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* **40**(5), 373-383.

Christensen K.B., Kreiner S. & Mesbah M. (Eds). (2013) *Rasch Models in Health*. London: Wiley.

Criteria Committee of the New York Heart Association. (1964) *Diseases of the heart and blood vessels: Nomenclature and criteria for diagnosis*, 6th ed. Boston, MA: Little Brown.

Gandaglia G., Briganti A., Jackson G., Kloner R.A., Montorsi F., Montorsi, P. & Vlachopoulos C. (2014) A systematic review of the association between erectile dysfunction and cardiovascular disease. *European Urolology* **65**(5), 968-978. doi: 10.1016/j.eururo.2013.08.023

Hagquist C., Bruce M. & Gustavsson P.J. (2009) Using the Rasch model in nursing: An introduction and illustrative example. *International Journal of Nursing Studies* **46**(3), 380-393. doi: org/10.1016/j.ijnurstu.2008.10.007

Hare D.L. & Davis C.R. (1996) Cardiac Depression Scale: validation of a new depression scale for cardiac patients**.** *Journal of Psychosomatic Research* **40**(4), 379-86. doi: org/10.1016/0022-3999(95)00612-5

Hawkins M.A.W., Dolansky M.A., Schaefer J.T., Fulcher M.J., Gunstad J., Redle J.D., Josephson R. & Hughes JW. (2015) Cognitive function in heart failure is associated with nonsomatic symptoms of depression but not somatic symptoms. *Journal of Cardiovascular Nursing* **30**(5), E9-E17. doi: 10.1097/JCN.0000000000000178

Hobart, J. & Cano, S. (2009) Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment* **13**(12), 1-177. doi: 10.3310/hta13120

Jonkman N.H., Schuurmans M.J., Groenwold R.H., Hoes A.W. & Trappenburg J.C. (2016) Identifying components of self-management interventions that improve health-related quality of life in chronically ill patients: Systematic review and meta-regression analysis. *Patient Education and Counseling*. **99**(7), 1087-1098. doi: 10.1016/j.pec.2016.01.022

Koski L., Xie H. & Finch L. (2009) Measuring cognition in a geriatric outpatient clinic: Rasch analysis of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology* **22**(3)**,** 151-160. doi: 10.1177/0891988709332944

Linacre, J.M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions* **7**(4), 328.

Lee C.S., Moser D.K., Lennie T.A. & Riegel, B. (2011) Event-free survival in adults with heart failure who engage in self-care management. *Heart and Lung* **40**(1), 12-20. doi: 10.1016/j.hrtlng.2009.12.003

Marais I. & Andrich D. (2008) Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement* **9**(3)**,** 200-215.

Masters G.N. (1982) A Rasch model for partial credit scoring. *Psychometrika* **47**(2), 149-74. doi: 10.1007/BF02296272

Nasreddine Z.S., Phillips N.A., Bedirian V., Charbonneau S., Whitehead V., Collin I., Cummings J.L. & Chertkow H. (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment**.** *Journal of the American Geriatric Society* **53**(4)**,** 695-699. doi: 10.1111/j.1532-5415.2005.53221.x

Newby V., Conner G., Grant C. & Bunderson V. (2010) *Rasch model and additive conjoint measurement*. In: M. Garner, G. Engelhard, M. Wilson & W. Fisher (Eds.) Advances in Rasch Measurement. Vol 1. Maple Grove, MN: JAM Press.

Rasch G. (1960) *Probabilistic models for some intelligence and attainment tests*. Copenhagen; Danish Institute for Educational Research. (Expanded Ed. 1980) Chicago, IL: University of Chicago Press.

Riegel B, Driscoll A., Suwanno J., Moser D.K., Lennie T.A., Chung M.L., Wu J.R., Dickson V.V., Carlson B. & Cameron J. (2009a) Heart failure self-care in developed and developing countries. *Journal of Cardiac Failure* **15**(6)**,** 508-516. doi: 10.1016/j.cardfail.2009.01.009

Riegel B., Lee C.S., Dickson V.V. & Carlson B. (2009b) An update on the Self-Care of Heart Failure Index. *Journal of Cardiovascular Nursing* **24**(6), 485-497. doi: 10.1097/JCN.0b013e3181b4baa0

Riegel B., Dickson V.V. & Faulkner K.M. (2016) The situation-specific theory of heart failure self-care: Revised and updated. *Journal of Cardiovascular Nursing* **31**(3)**,** 226-235. doi: 10.1097/JCN.0000000000000244

Scheuneman J.D. & Subhiyah R.G. (1998) Evidence for the validity of a Rasch model technique for identifying differential item functioning. *Journal of Outcome Measurement* **2**(1), 33-42.

Smith A.B., Rush R., Fallowfield L.J., Velikova G. & Sharpe, M. (2008) Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology* **8**, 33. doi: 10.1186/1471-2288-8-33

Smith E.V. (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement* **3**(2), 205-231.

Sokoreli I., de Vries J.J., Pauws S.C. & Steyerberg E.W. (2016) Depression and anxiety as predictors of mortality among heart failure patients: systematic review and meta-analysis. *Heart Failure Reviews* **21**(1), 49-63. doi: 10.1007/s10741-015-9517-4

Tennant A. & Conaghan P.G. (2007a) The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied and what should one look for in a Rasch paper? *Arthritis and Rheumatology* **57**(8), 1358-1362. doi: 10.1002/art.23108

Vellone E., Jaarsma T., Strömberg A., Fida R., Arestedt K., Rocco G., Cocchieri A. & Alvaro R. (2014) The European Heart Failure Self-care Behaviour Scale: new insights into factorial structure, reliability, precision and scoring procedure. *Patient Education and Counseling* **94**(1)**,** 97-102. doi: 10.1016/j.pec.2013.09.014

World Medical Association. (2010) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Journal of the American Medical Association* **310**(20), 2191-2194. doi: 10.1001/jama.2013.281053.

Wright B.D. (1996) Local dependency, correlations and principal components. *Rasch Measurement Transactions* **10**(3), 509-511.

Wright B.D. (2001) Separation, reliability and skewed distributions: statistically different levels of performance. *Rasch Measurement Transactions* **14**(4), 786.

Yen W.M. (1993) Scaling performance assessments - Strategies for managing local item dependence. *Journal of Educational Measurement* **30**, 187-213. doi: 10.1111/j.1745-3984.1993.tb00423.x

TABLES

Table 1

Heart failure participant characteristics

Variable	**N**	**Sample**
Age, Median (range) years	135	68 (28 - 89)
Gender, male	135	95 (70%)
English first language	133	110 (83%)
Married/de facto relationship	135	71 (53%)
Years of education, Median (range)	131	12 (3 - 21)
Diagnosis of HF <2 months	131	23 (17%)
Ischaemic aetiology of HF	124	69 (56%)
NYHA classification	135	
I – II		91 (67%)
III – IV		44 (33%)
Charlson Comorbidity Index, Median (range)	135	3 (1 to 10)

NYHA = New York Heart Association

Table 2

Summary fit statistics following each modification

			Persons fit residuals		Item fit residuals		Item-trait interaction			Internal consistency
Stage		Modification	$\bar{X}$	*SD*	$\bar{X}$	*SD*	χ^2	*df*	*p*	(PSI)*
		All modifications	**-0.05**	**(0.89)**	**0.05**	**(0.73)**	**39.45**	**42**	**.61**	**.61**
0	Initial fit	No modifications	-0.06	(0.98)	0.18	(0.99)	95.50	54	.0004	.69
1	Threshold disorder	Re-coded rating scales	-0.02	(0.70)	0.01	(1.03)	85.06	54	.004	.59
2	Poor discrimination	Deleted items	-0.05	(0.75)	-0.05	(0.78)	52.67	48	.30	.62
3	Local dependence	Amalgamated or deleted items	-0.05	(0.74)	-0.04	(0.89)	47.46	44	.33	.62
4	Differential item functioning	Deleted item	-0.06	(0.75)	0.07	(0.91)	42.54	42	.45	.61

* the person separation index (PSI), a measure of the internal consistency reliability of measures

Data shows standardized fit residual distribution means ($\bar{X}$) and standard deviations (SD) for persons and items and the item-trait interaction chi-square χ^2 and degrees of freedom (df)

Table 3 Individual item fit statistics, parameter estimates and standard errors of the modified unidimensional Heart-FaST

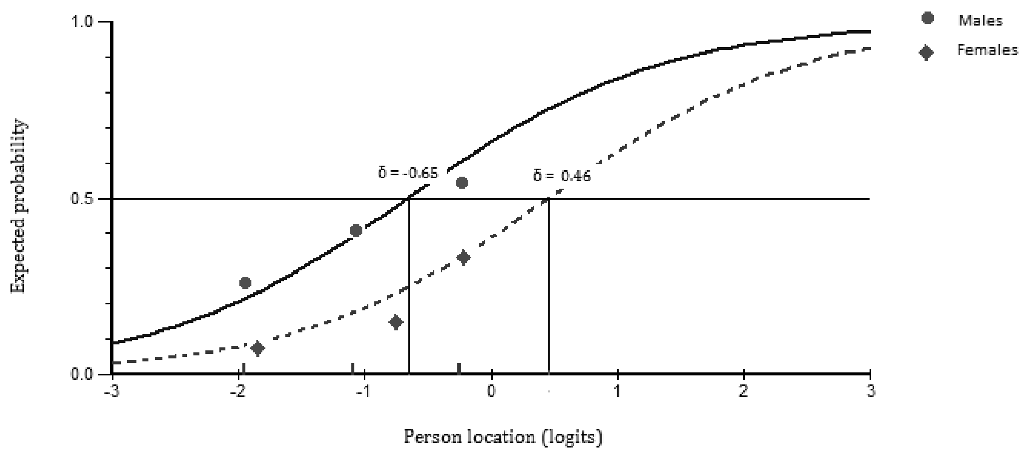
		Location		Fit statistics		
		(δ)	*SE*	Residual	χ^2	*p*
PFD	Have someone to confide in	1.56	(0.34)	-0.38	1.13	0.57
CF	Clock Contours	1.14	(0.29)	-0.45	0.55	0.75
CF	Serial Sevens	0.93	(0.18)	-0.64	3.22	0.19
EF	Only misery in the future	0.77	(0.25)	0.04	1.14	0.56
CF	Verbal Fluency: Animals	0.63	(0.17)	-0.61	0.43	0.80
PD	NYHA functional classes	0.61	(0.17)	-0.83	2.39	0.30
PFD	Not fluent in English	0.44	(0.23)	-0.91	1.96	0.37
CF	Had difficulty keeping attention on an activity for long	0.14	(0.21)	-0.82	1.95	0.38
EF	Feel in good spirits	0.12	(0.21)	-1.22	5.38	0.06
CF	Clock Numbers	0.08	(0.21)	1.02	1.20	0.12
EF	Feel independent and in control	-0.04	(0.21)	-0.62	2.03	0.36

CF	Verbal fluency: Letters	-0.08	(0.14)	-0.06	1.56	0.45
CF	Digit Span Task	-0.22	(0.14)	-0.76	0.73	0.69
CF	Had others notice or comment on trouble with memory	-0.31	(0.20)	-0.01	0.29	0.86
EF	Had life experiences that continue to bother me	-0.32	(0.20)	1.22	1.54	0.46
EF	Concentration is good	-0.57	(0.19)	-1.66	6.32	0.04
CF	Clock Hands	-0.71	(0.19)	1.33	0.38	0.82
PFD	Over 70 years of age	-0.94	(0.19)	-1.09	0.53	0.77
PFD	Charlson comorbidity index	-0.96	(0.12)	1.36	5.44	0.07
EF	Living on borrowed time	-1.05	(0.18)	0.45	0.31	0.86
CF	Had trouble with memory	-1.22	(0.18)	0.72	1.04	0.59

PFD: Physical functioning & demographic risk factors. CF: Cognitive Functioning. EF: Emotional Functioning; CDT = Clock Drawing Task. Bonferroni adjusted significance level of 0.002. Chi-squared tests based on three class intervals comprising approximately equal numbers of participants. Chi-Squared degrees of freedom = 2.

Individual item fit statistics, calibrations, and standard errors following modifications to heart fast. Items are presented in order of difficulty (δ) from relatively high to low risk of poor self-care.

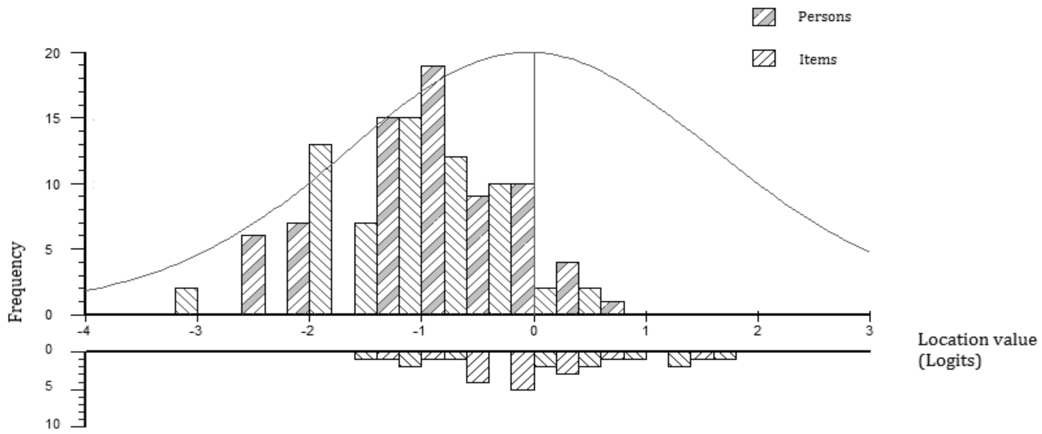
FIGURES



Note: Class intervals: Male 31, 30, 33; Female 16, 15, 9.

Response probabilities (sigmoid function) and observed proportions are plotted separately for males and females as a function of risk.

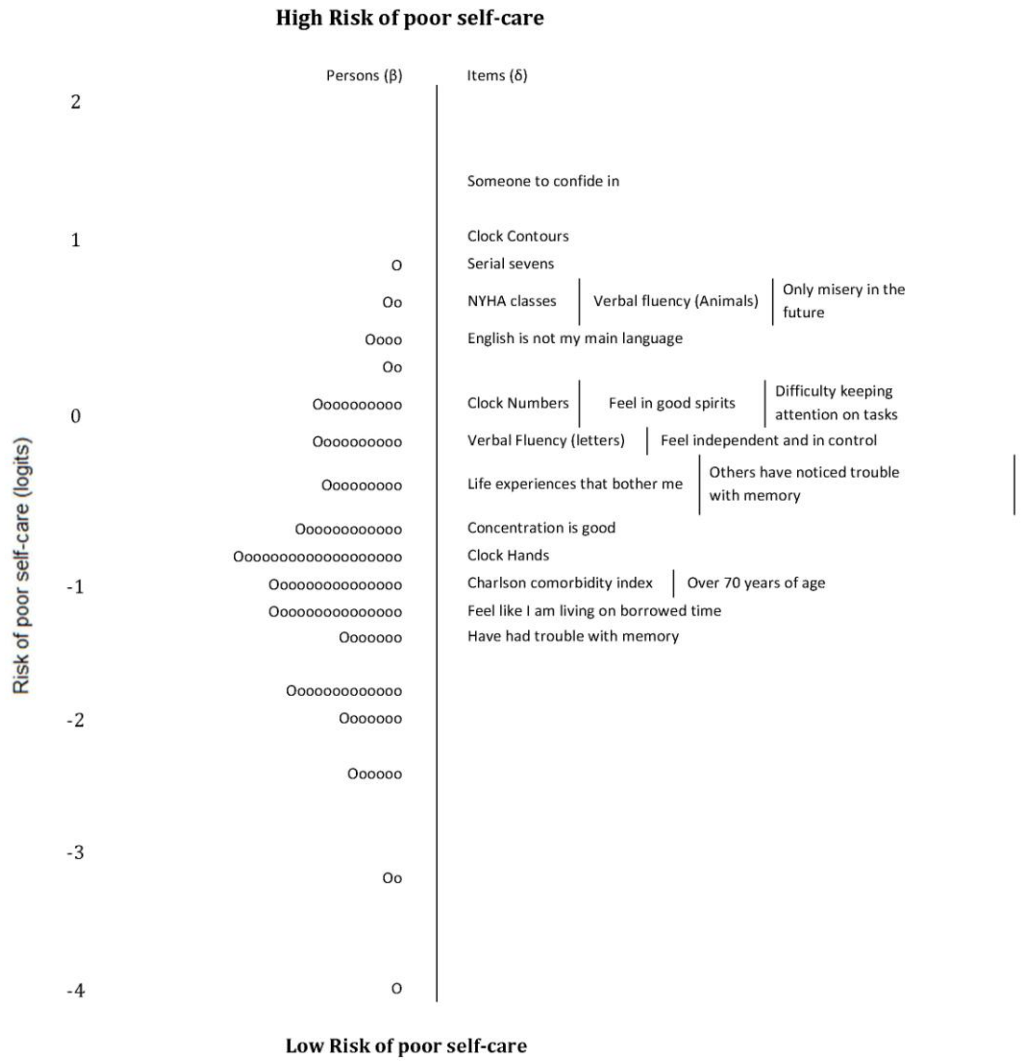
Fig. 1. Differential item functioning in relation to the self-report item, ‘concern regarding my capacity for sexual activity’



The bottom bars show the locations of dichotomous items and the threshold locations of polytomous items. The top bars show the locations of HF patients.

The alignment of persons and items indicates that Heart-FaST optimally targets people located in the higher region of poor self-care risk. The information function (superimposed on the top bars) peaks where measures of risk are most precise, at approximately zero logits, but more than one logit above the average person measure in this sample.

Fig. 2. The relative locations of items (or thresholds) and persons along the continuum of risk



Persons and items scaled in a common, log-odds unit along the continuum of poor self-care risk. The right hand side shows Heart-FaST items in order of difficulty while the left hand side shows the location of persons.

Fig.3. Item person map